

## REMARKS/ARGUMENTS

Claims 1-72 were previously pending in the application. Claims 27, 33, 37-38, 49, 52, 55, 59-60, 63-64, and 68 are amended; and new claims 73-77 are added herein. Assuming the entry of this amendment, claims 1-77 are now pending in the application. The Applicant hereby requests further examination and reconsideration of the application in view of the foregoing amendments and these remarks.

### Miscellaneous Claim Amendments

Claims 55, 63, and 64 have been amended to refer to the number of playback channels generated from the  $E$  transmitted channels using the parameter  $M$ , instead of the parameter  $C$ , in order to clarify that the number of playback channels can be (but does not need to be) different from the number of input channels (i.e.,  $C$ ) recited in claims 27, 37, and 38 as being used to generate the  $E$  transmitted channels. Claims 63 and 64 have also been amended to clarify that  $M \geq E \geq 1$ , as recited in previously presented claim 55. Claims 59 and 68 have been amended to conform to the amendments to claims 55 and 64, respectively. Support for these amendments is found on page 19, lines 5-12, of the specification. The Applicant submits that none of these amendments were made to overcome any prior-art rejections.

### Claim Objections

In paragraph 1 of the office action, the Examiner objected to claims 33 and 60. In response, the Applicant has amended claims 33 and 60 as suggested by the Examiner. The Applicant submits that these amendments were not made to overcome any prior-art rejections.

### Drawings

In paragraph 2, the Examiner stated that Fig. 11 should be designated as "Prior Art." In response, the Applicant submits herewith a Transmittal of Corrected Drawing(s) amending Fig. 11 as suggested by the Examiner.

In paragraph 3, the Examiner objected to the drawings under 37 CFR 1.83(a), implying that the drawings do not show the claimed features that are related to downmixing and upmixing. As known in the art, in the context of audio processing, downmixing refers generally to the processing of a number of input audio channels to generate a smaller number of output audio channels, while upmixing refers generally to the processing of a number of input audio channels to generate a larger number of output audio channels.

With this understanding of the terms "downmixing" and "upmixing," the Applicant submits that the drawings do show the claimed features that are related to downmixing and upmixing. For example, as shown in Fig. 4, combiner 404 receives  $N$  input audio channels (i.e., source signals 1, 2, ...,  $N$ ) and processes those input audio channels to generate a smaller number of output audio channels (i.e., the combined signal). As such, combiner 404 performs downmixing. Downmixing is also performed by auditory scene removal 1006 of Fig. 10, PCSC encoder 1201 of Figs. 12 and 13, PCSC encoder 1401 of Fig. 14, and PCSC encoder 1501 of Fig. 15.

Similarly, as shown in Fig. 7, auditory scene synthesis 704 receives a single input audio channel (from TF transform 702) and processes that input audio channel to generate a larger number of output audio channels (i.e., the two audio channels applied to inverse TF transforms 706). As such, auditory

scene synthesis 704 performs upmixing. Upmixing is also performed by PCSC decoder 1209 of Fig. 12, PCSC decoder 1409 of Fig. 14, and PCSC decoder 1509 of Fig. 15.

In view of the foregoing, the Applicant submits that the drawings do show the claimed features that are related to downmixing and upmixing and that the Examiner's objection to the drawings under 37 CFR 1.83(a) has been overcome.

### Specification

In paragraph 4, the Examiner objected to the disclosure because "There is no mention in the specification regarding downmixing or downmixer and upmixing or upmixer." For the reasons provided in the previous section, the Applicant submits that the disclosure does adequately disclose the concepts of downmixing and upmixing "to allow one of ordinary skill [to] make or use the claimed invention" and that the Examiner's objection to the disclosure has been overcome.

### Claim Rejections

In paragraph 6, the Examiner rejected claims 1-10, 12-23, and 25-72 under 35 U.S.C. 103(a) as being unpatentable over Ten Kate in view of Shaffer, and further in view of Moon. In paragraph 7, the Examiner rejected claims 11 and 24 under 35 U.S.C. 103(a) as being unpatentable over Ten Kate in view of Shaffer, further in view of Moon, and further in view of Jafarkhani. For the following reasons, the Applicant submits that all of the now-pending claims are allowable over the cited references.

#### Claims 27, 37, 38, 49, and 52

Claim 27 has been amended to clarify that an audio decoder is enabled to generate more than  $E$  different playback audio channels based on the  $E$  transmitted channels and the one or more cue codes. For example, if  $E=1$ , then an audio encoder is enabled to generate two or more different playback audio channels based on the single transmitted channel and the one or more cue codes. Support for the amendment to claim 27 is found, for example, in the last paragraph of previously presented claim 28.

Ten Kate teaches N-channel transmission that is compatible with 2-channel transmission and 1-channel transmission. See Title of the Invention. For example, Figs. 1-15 teach a 3-channel transmission system in which three input audio channels L, C, and R are used to generate three transmitted audio channels  $M_o$ , AUX1, and AUX2. As indicated in Fig. 9b, a 3-channel receiver that receives all three transmitted channels (170) can generate three playback audio channels L, C, and R (171). As shown in Fig. 11, a prior-art 2-channel decoder apparatus ignores AUX2 and processes  $M_o$  and AUX1 to generate two playback audio channels. See column 11, lines 39-65. As shown in Fig. 12, a prior-art mono decoder apparatus ignores AUX1 and AUX2 and processes  $M_o$  to generate one playback channel. See column 11, line 66, to column 12, line 18.

As indicated in Fig. 9b and as described in column 10, line 66, to column 11, line 9, when only two channels ( $M_o$  and AUX1) are transmitted (172), a decoder apparatus can generate at most two playback channels (173). Similarly, as described in column 11, lines 10-18, when only one channel ( $M_o$ ) is transmitted (174), a decoder apparatus can generate only one playback channel (175), which Fig. 9b shows being applied to each of three loudspeakers.

Significantly, in Ten Kate, the number of different playback channels is never greater than the number of transmitted channels. Three transmitted channels can be used to generate 1, 2, or 3 playback channels, but Ten Kate does not teach or even suggest (i) generating more than two different playback

channels when only two channels are transmitted or (ii) generating more than one different playback channel when only one channel is transmitted.

In rejecting previously presented claim 28, the Examiner suggested that Ten Kate teaches "the format enables a second audio decoder (standard stereo decoder, col. 2, lines 50-56) having knowledge of the existence of the one or more cue codes (in combination signal) to generate more than E playback audio channels based on the E transmitted channels and the one or more cue codes," citing column 2, lines 50-59, and column 5, lines 17-24. The Applicant submits that the Examiner mischaracterized the teachings in Ten Kate in rejecting previously presented claim 28.

First of all, there are no cue codes in Ten Kate's combination signal, as the Examiner admitted on page 5 ("Ten Kate does not explicitly disclose data stream including cue codes.")

Moreover, in column 2, lines 50-59, Ten Kate teaches generating 1, 2, or 3 playback channels from three transmitted channels. Similarly, column 5, lines 17-24, teaches that a mono decoder generates one playback channel from the three transmitted channels, while a standard stereo decoder generates two playback channels from the three transmitted channels.

Nowhere does Ten Kate teach or even suggest generating a greater number of different playback channels than the number of transmitted channels. As such, the Examiner mischaracterized the teachings in Ten Kate in rejecting previously presented claim 28.

According to claim 27, two or more of the *C* input channels are provided in a frequency domain, and one or more cue codes are generated for each of one or more different frequency bands in the two or more input channels in the frequency domain. Since Ten Kate does not teach the generation of cue codes, Ten Kate cannot possibly be said to teach the generation of cue codes "for each of one or more different frequency bands in the two or more input channels in the frequency domain."

Nor do the other cited references teach these features. For example, in Shaffer, interaural time delays are generated in the time domain. See, e.g., Figs. 3, 5, and 6; and column 6, line 61, to column 9, line 44. Shaffer does not teach or even suggest any techniques for generating interaural time delays in anything but the time domain. (Note that the "subband coding" mentioned in column 3, line 45, refers to the conventional audio coding applied to Shaffer's single sample stream generated by adder 44 (see Fig. 3 and column 6, lines 18-26), not to the generation of Shaffer's cue codes.)

Moon does not teach or suggest the features of the present invention that are missing from Ten Kate and Shaffer.

For all these reasons, the Applicant submits that currently amended claim 27 is allowable over the cited references. For similar reasons, the Applicant submits that currently amended claims 37, 38, 49, and 52 are allowable over the cited references. Since the rest of the claims depend variously from claims 27, 37, 38, 49, and 52, it is further submitted that those claims are also allowable over the cited references.

#### Claims 35 and 46

According to claim 35, the downmixing comprises, for each of one or more different frequency bands, downmixing the two or more input channels in the frequency domain into one or more downmixed channels in the frequency domain. In rejecting claim 35, the Examiner suggested that

Shaffer teaches downmixing in the frequency domain, stating that "the decoder can apply the balance parameter to all received frequencies" and citing column 5, lines 56-67.

First of all, whether or not Shaffer's decoder can apply the balance parameter to all received frequencies is completely irrelevant to whether or not the downmixing is performed in the frequency domain. In Shaffer (and in the present invention), downmixing is performed at the encoder, not at the decoder. See, e.g., adder 44 in Shaffer's Fig. 3. Significantly, the only downmixing taught in Shaffer is performed in the time domain, not in the frequency domain. See, e.g., column 6, lines 18-26.

There is no teaching or even suggestion in Shaffer for performing downmixing in the frequency domain. As such, the Applicant submits that this provides additional reasons for the allowability of claim 35 and also claim 46 over the cited references.

#### Claims 36 and 47

According to claim 36, the downmixing further comprises converting the one or more downmixed channels from the frequency domain into one or more of the transmitted channels in the time domain. Thus, according to claim 36, which depends from claim 35, after the downmixed channels are generated in the frequency domain, they are converted into transmitted channels in the time domain. In rejecting claim 36, the Examiner suggested that "time shifting" is related to converting downmixed channels in the frequency domain into transmitted channels in the time domain, citing column 6, lines 18-26, of Shaffer.

First of all, the time shifting mentioned in column 6, lines 23-24, refers to a technique for generating a single sample stream from left and right sample streams in which one of the input sample streams is shifted in time relative to the other input sample stream to compensate for a characterized time delay between the two input streams before the input streams are combined to generate the single sample stream. This processing is implemented entirely in the time domain and has absolutely nothing to do with converting audio channels from a frequency domain into a time domain.

Moreover, this time shifting is performed before the two input streams are combined to generate the single (downmixed) sample stream. As such, this time shifting is performed before the downmixed sample stream is even generated.

In fact, there is no teaching or even suggestion in Shaffer for converting downmixed channels from a frequency domain into a time domain. The Applicant submits that this provides additional reasons for the allowability of claims 36 and 47 over the cited references.

#### Claims 3 and 16

According to claim 3, each set of one or more auditory scene parameters corresponds to a different audio source in the auditory scene. In rejecting claim 3, the Examiner cited Shaffer's interaural level differences (ILD) and interaural time delays (ITD) and column 4, lines 38-44. While it is true that Shaffer's ILD and ITD are two different auditory scene parameters, the number of different auditory scene parameters has nothing to do with the number of audio sources in an auditory scene for which the different auditory scene parameters are derived. An audio source in an auditory scene refers to the location in physical space of the origin of sound arriving at the microphones that are used to generate the auditory scene parameters.

According to claim 2, from which claim 3 depends, each set of auditory scene parameters corresponds to a different frequency band in the combined audio signal. Claim 3 adds the limitation that each set of auditory scene parameters corresponds to a different audio source in the auditory scene. Thus, if the sound arriving at the microphones in an auditory scene originates at multiple locations (e.g., when someone is speaking in a conference room having an air conditioner running and a car passing by outside), according to claim 3, each set of auditory scene parameters corresponds to a different frequency band (e.g., one set of auditory scene parameters in a first frequency band corresponds to the speaker, another set in a second frequency band corresponds to the air conditioner, and yet another set in a third frequency band corresponds to the car).

Clearly, Shaffer does not teach or even suggest the features recited in claim 3. The Applicant submits that this provides additional reasons for the allowability of claim 3 over the cited references. Similarly, the Applicant submits that this provides additional reasons for the allowability of claim 16.

#### Claims 4 and 17

According to claim 4, for at least one of the sets of one or more auditory scene parameters, at least one of the auditory scene parameters corresponds to a combination of two or more different audio sources in the auditory scene that takes into account relative dominance of the two or more different audio sources in the auditory scene. In rejecting claim 4, the Examiner cited Shaffer's use of cross-correlation, Fig. 7, and column 8, line 43-57. The Applicant submits that Shaffer's use of cross-correlation has nothing to do with the features recited in claim 4.

According to claim 4, at least one auditory scene parameter corresponds to sound coming from two or more different locations in the auditory scene, where the relative dominance (e.g., which location is louder than the other(s)) of the audio sources is taken into account. Shaffer uses cross-correlation between two different audio channels to determine the relative time delay between the two channels. This has nothing to do with the relative dominance of the two audio channels. Moreover, Shaffer's two audio channels are left and right stereo channels, which is independent of whether the sound in those stereo audio channels comes from one audio source or multiple audio sources.

Clearly, Shaffer does not teach or even suggest the features recited in claim 4. The Applicant submits that this provides additional reasons for the allowability of claim 4 over the cited references. Similarly, the Applicant submits that this provides additional reasons for the allowability of claim 17.

#### Claims 7 and 20

According to claim 7, the combined audio signal corresponds to a combination of two or more different mono source signals, wherein the two or more different frequency bands are selected by comparing magnitudes of the two or more different mono source signals, wherein, for each of the two or more different frequency bands, one of the mono source signals dominates the one or more other mono source signals. In rejecting claim 7, the Examiner cited both Ten Kate and Shaffer.

In particular, the Examiner suggested that Ten Kate teaches a combined audio signal corresponding to a combination of two or more different mono source signals, citing column 2, lines 33-56. Ten Kate teaches a combined audio signal corresponding to a combination of three audio signals, but they are not three different mono source signals. Three different mono source signals refers to three different mono audio signals, each coming from a different location in an auditory scene. In Ten Kate, the three audio signals are the left (L), right (R), and center (C) channels of a 3-channel audio system, where the sound in those three different channels all come from the same audio source(s) in an auditory

scene. Thus, Ten Kate does not teach or even suggest a combined audio signal corresponding to a combination of two or more different mono source signals.

The Examiner also suggested that Shaffer teaches that "for each of the two or more different frequency bands, one of the mono source signals dominates the one or more other mono source signals," again citing Shaffer's use of cross-correlation. For at least some of the same reasons given in the previous section, the Applicant submits that Shaffer's use of cross-correlation has nothing to do with relative dominance of audio sources. Moreover, Shaffer uses cross-correlation to determine relative time delays between audio channels. There is no teaching or even suggestion in Shaffer for using cross-correlation to select different frequency bands.

The Applicant submits that this provides additional reasons for the allowability of claim 7 over the cited references. Similarly, the Applicant submits that this provides additional reasons for the allowability of claim 20.

#### Claims 8 and 21

According to claim 8, the combined audio signal corresponds to a combination of left and right audio signals of a binaural signal, wherein each different set of one or more auditory scene parameters is generated by comparing the left and right audio signals in a corresponding frequency band. In rejecting claim 8, the Examiner suggested that Shaffer teaches generating auditory scene parameters "by comparing the left and right audio signals in ... corresponding frequency bands," citing Shaffer's "subband coding" and column 3, lines 43-47.

First of all, Shaffer teaches the generation of ITD parameters using only time-domain techniques. There is no teaching in Shaffer for generating ITD parameters in the frequency domain, let alone generating different ITD parameters in different frequency bands.

Moreover, the subband coding described in Shaffer is performed by signal encoder 46 of Fig. 4 on the single sample stream generated by adder 44. Since, at this point, there is only one audio channel, there simply cannot be any comparison between left and right audio signals in Shaffer's subband coding.

The Applicant submits that this provides additional reasons for the allowability of claim 8 over the cited references. Similarly, the Applicant submits that this provides additional reasons for the allowability of claim 21.

#### Claims 10 and 23

According to claim 10, step (b) comprises the step of applying a layered coding technique in which stronger error protection is provided to the combined audio signal than to the auditory scene parameters when generating the embedded audio signal, such that errors due to transmission over a lossy channel will tend to affect the auditory scene parameters before affecting the combined audio signal to improve the probability of the first receiver to process at least the combined audio signal.

In rejecting claim 10, the Examiner suggested that Shaffer teaches such a layered coding technique, citing column 3, lines 63-67; column 1, lines 21-24; voice packets 50 in Fig. 4; encoder 24 and decoder 30 of Fig. 1; Fig. 7; and column 8, lines 43-50. The Applicant submits that the Examiner mischaracterized the teachings in Shaffer in rejecting claim 10.

Column 3, lines 63-67, describes packets being encapsulated with "lower layer headers." Column 1, lines 21-24, suggests that packet headers can contain "error correction information." Fig. 4 indicates the generation of "voice packets 50." These teachings suggest, at most, that, by using voice packets with headers that contain error correction information, Shaffer's encoder 24 and decoder 30 work together to recover from some errors due to transmission over a lossy channel.

Significantly, however, there is no teaching or suggestion in Shaffer of a layered coding technique in which stronger error protection is provided to the combined audio signal than to the auditory scene parameters when generating the embedded audio signal, such that errors due to transmission over a lossy channel will tend to affect the auditory scene parameters before affecting the combined audio signal to improve the probability of the first receiver to process at least the combined audio signal. The Examiner's bald assertion that such features are obvious in light of Shaffer's teachings is completely unsupported by any teachings in Shaffer and therefore improper.

Note that Fig. 7 and column 8, lines 43-50, relate to cross-correlation processing used to determine the time delay between the left and right audio channels. These teachings have absolutely nothing to do with error protection.

The Applicant submits that this provides additional reasons for the allowability of claim 10 over the cited references. Similarly, the Applicant submits that this provides additional reasons for the allowability of claim 23.

#### Claim 55 and 63

According to currently amended claim 55, for each of one or more different frequency bands, one or more of the  $E$  transmitted channels are upmixed in a frequency domain to generate two or more of  $M$  playback channels in the frequency domain, where  $M > E \geq 1$ . The one or more cue codes are applied to each of the one or more different frequency bands in the two or more playback channels in the frequency domain to generate two or more modified channels, and the two or more modified channels are converted from the frequency domain into a time domain.

In rejecting claim 55, the Examiner appears to have combined different passages in Ten Kate, Shaffer, and Moon based on the appearance of individual words that are also recited in claim 55, with little or no regard for the teachings in those passages or the lack of motivation for such combinations.

For example, the Examiner cited Moon, column 7, lines 42-57, as being related to the upmixing of claim 55. According to claim 55, upmixing is applied to one or more transmitted channels to generate two or more playback channels. In this context, upmixing is related to a technique for increasing the number of channels by variously duplicating and/or combining the input channels. Moon teaches a completely different type of upmixing. The upmixing taught in Moon relates to the conversion of a single input channel in one frequency into a single output channel of a higher frequency. In this type of upmixing, the number of channels does not change; only the channel frequency changes.

Similarly, the Examiner cited the "subband coding" mentioned in Ten Kate, column 6, lines 47-59, as being related to the fact that the upmixing of claim 55 is performed in the frequency domain. Like the subband coding taught in Shaffer described previously, the "subband coding" taught in Ten Kate refers to the conventional data compression technique applied to a single sample stream. It has nothing to do with upmixing one or more input channels to generate a greater number of output channels.

Here, too, the Examiner cited Ten Kate, column 2, lines 28-38, as teaching the generation of more playback channels than the number of transmitted channels. As described previously with regard to claim 27, Ten Kate does not teach or even suggest the generation of more playback channels than the number of transmitted channels.

The Examiner cited Shaffer, column 1, line 60, to column 2, line 7, and Figs. 4 and 8 as being related to the application of cue codes to different frequency bands in the playback channels in the frequency domain to generate two or more modified channels, recited in claim 55. The passage cited by the Examiner teaches the application of a directional cue, but there is no teaching or suggestion that this processing is performed on different frequency bands in the frequency domain. Fig. 4 shows a packet format, and Fig. 8 shows Shaffer's decoder. Neither of these figures shows or suggests any frequency-domain processing.

Regarding the conversion of the modified channels from the frequency domain into a time domain, the Examiner cited Shaffer, column 4, lines 11-22. The teachings in this passage relate to the handling of transmitted data packets and has absolutely nothing to do with the conversion of channels from a frequency domain into a time domain.

Even if the Examiner were correct about the individual teachings in the various references (which the Applicant explicitly and emphatically denies), the fact remains that there is no motivation for such a combination of references. An Examiner is not free to haphazardly combine references from disparate references to reject a claimed invention. There has to be a legitimate motivation for such a combination.

The Applicant submits that this provides additional reasons for the allowability of claim 55 and similarly of claim 63 over the cited references.

#### Claims 62 and 71

According to claim 62, the upmixing comprises, for each of one or more different frequency bands, upmixing at least two of the  $E$  transmitted channels into at least one playback channel in the frequency domain. In rejecting claim 62, the Examiner cited Shaffer, column 5, lines 56-67, as teaching "the upmixing comprises, for each of one or more different frequency bands, downmixing the two or more input channels in the frequency domain into one or more downmixed channels in the frequency domain" (emphasis added), stating further that "the decoder can apply the balance parameter to all received frequencies."

First of all, the recitations of claim 62 relate to upmixing at the decoder. The Examiner mischaracterized the recitations of claim 62 as "the upmixing comprises ... downmixing."

Furthermore, as described previously with regard to claim 35, whether or not Shaffer's decoder can apply the balance parameter to all received frequencies has nothing to do with whether or not upmixing is performed in the frequency domain. There is simply no teaching in Shaffer that any upmixing is performed in the frequency domain.



New Claims 73-77

According to new claim 73, the method comprises generating, in the frequency domain, ICTD data as one of the one or more cue codes. None of the cited references teaches generating ICTD data in a frequency domain. As such, the Applicants submits that this provides additional reasons for the allowability of claim 73 and similarly of claims 74-77 over the cited references.

In view of the foregoing, the Applicant respectfully submits that the rejections of claims under Section 103(a) have been overcome.

In view of the above amendments and remarks, the Applicant believes that the now-pending claims are in condition for allowance. Therefore, the Applicant believes that the entire application is now in condition for allowance, and early and favorable action is respectfully solicited.

Respectfully submitted,

Date: 10/3/06  
Customer No. 46900  
Mendelsohn & Associates, P.C.  
1500 John F. Kennedy Blvd., Suite 405  
Philadelphia, Pennsylvania 19102

Steve Mendelsohn  
Steve Mendelsohn  
Registration No. 35,951  
Attorney for Applicant  
(215) 557-6657 (phone)  
(215) 557-8477 (fax)